DOCUMENT RESUME

ED 090 510 CS 001 063

AUTHOR	Van Mondfrans, Adrian P.
TITLE	Learning Principles and Procedures Applied to Remedial Reading with Normal and Educable Mentally Retarded Fourth, Fifth, and Sixth Grade Children.
INSTITUTION	Wisconsin Univ., Madison. Research and Development Center for Cognitive Learning.
SPONS AGENCY	Office of Education (DHEW), Washington, D.C.
REPORT NO	TR-62
PUB DATE	Sep 68
CONTRACT	OEC-5-10-154
NOTE	31p.; Report from Project Motivated Learning
EDRS PRICE	MF-$0.75 HC-$1.85 PLUS POSTAGE
DESCRIPTORS	*Educable Mentally Handicapped; Intermediate Grades; Reading; Reading Achievement; *Reading Research; Reading Skills; *Remedial Reading Programs
IDENTIFIERS	*Staats Remedial Reading Procedures

ABSTRACT

This study was designed to further validate the Staats remedial reading procedures and to demonstrate their generality by extending their application to a population of pre- and early-adolescent retarded readers. The subjects consisted of 32 students from fourth, fifth, and sixth grade classrooms and 16 subjects of equivalent chronological age from special classes for the mentally retarded. The subjects were assigned to eight different treatment groups. The eight groups were defined by all possible combinations of three dichotomous factors--type of subject, type of reinforcement, and treatment. In general, the 24 experimental subjects participated in a manner substantiating the motivational effects of the extrinsic reinforcers. The mean number of reading responses made during the experiment was over 51,000. A mean of 550 new (unknown) words was presented with an average retention rate of 79 percent of these words on a short-term retention test and 54 percent on a long-term test. The results indicated that the experimental subjects scored significantly higher on a 100-item word recognition task and made a smaller percentage of errors in a prose reading task than did the control subjects. (WR)

U.S. DEPARTMENT OF HEALTH,
EDUCATION & WELFARE
NATIONAL INSTITUTE OF
EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

ED 090510

Technical Report No. 62

LEARNING PRINCIPLES AND PROCEDURES APPLIED TO REMEDIAL READING WITH NORMAL AND EDUCABLE MENTALLY RETARDED FOURTH, FIFTH, AND SIXTH GRADE CHILDREN

By Adrian P. Van Mondfrans

Report from Project Motivated Learning
Arthur W. Staats, Principal Investigator
(now at the University of Hawaii)

Wisconsin Research and Development
Center for Cognitive Learning
The University of Wisconsin
Madison, Wisconsin

September 1968

The author gratefully acknowledges the direction and support of A. W. Staats.

The author is now Assistant Professor of Education and Psychology, Purdue University. The research reported herein was performed pursuant to a contract with the United States Office of Education, Department of Health, Education and Welfare, under the provisions of the Cooperative Research Program.

Center No. C-03 / Contract OE 5-10-154

CS 001 063

PREFACE

A goal of this Center is to conduct both laboratory and classroom research in order to create knowledge, theory, and methods which will effectively improve learning in our schools. One research project, directed by Professor Staats, has been concerned with the development of effective training procedures. From this project comes the present report of Mr. Van Mondfrans' attempt to apply these procedures to the problem of training educable retarded school children to read.

It is interesting to note not only the success of these training methods but that these procedures can be used quite effectively by nonprofessional personnel.

Harold J. Fletcher
Director of Program 1

CONTENTS

		page
	List of Tables and Figures	vii
	Abstract	ix
I.	Introduction	1
II.	Method and Procedures	6
	Selection and Assignment of Subjects	6
	Special Problems with Subjects	6
	Instruments	8
	Instructional Technicians	8
	Description of Instructional Materials	9
	Procedures	9
	Individual Word Phase	9
	Oral Reading Phase	10
	Silent Reading Phase	10
	Comprehension Phase	10
	Vocabulary Review	10
	The Token System	11
	Bonus System	11
	The Session	11
III.	Results	13
	The Response Data Internal to the Program	13
	The Total Number of Reading Responses	13
	Words Known upon First Presentation and on Retention Tests	14
	The Amount of Reinforcement Per Reading Response	15
	Results of Test Data	15
	The Lorge-Thorndike Group Intelligence Test	15
	The California Reading Test	15
	100-WRT Test Results	15
	The 5-Minute Story Test	16
IV.	Discussion and Conclusions	
	Test Results	20
	The Standardized Test Results	20
	The Nonstandardized Test Results	20
	Summary of the Test Data	21
	The Effect of Type of Reinforcement on the Test Data	21

LIST OF TABLES AND FIGURES

Table		page
1	Breakdown of Ss by Groups Within Schools	7
2	Means and Standard Deviations of Pretest Scores by Variable and Groups	7
3	Analysis of Covariance Summary Table with Total Reading Responses as Criterion and 100-WRT Pretest Scores as the Covariate	13
4	Means and Standard Deviations of Posttest Scores and Gain Scores by Variable and Groups	17
5	Summary Table for Analysis of Covariance with Posttest DIQ Scores on the Lorge-Thorndike Group Intelligence Test as Criterion and Pretest DIQ Scores as Covariate	17
6	Analysis of Covariance Summary Table with Posttest California Reading Test, Total Reading, as Criterion and California Reading Test, Total Reading, Pretest as Covariate	17
7	Analysis of Covariance Summary Table with 100-WRT Posttest Scores as the Criterion Variable and 100-WRT Pretest Scores as the Covariate	18
8	Analysis of Covariance Summary Table with Words/Minute Scores from the 5-Minute Story Test as Criterion and Pretest Scores on the 100-WRT as the Covariate	18
9	Summary Table for Analysis of Covariance with %-Errors on Five-Minute Story Test as Criterion and Pretest Scores on the 100-WRT Test as Covariate	18

Figure		
1	The Cumulative Mean Number of Reading Responses as a Function of Sessions Completed for the Four Experimental Groups	14
2	The Payrate as a Function of Sessions Completed for the Four Experimental Groups	16

ABSTRACT

Arthur W. Staats has devised remedial reading procedures based upon traditional conditioning principles (Staats & Butterfield, 1965; Staats, Minke, Goodwin, & Landeen, 1967). The procedures were sufficiently simple and easy to administer that, in the two studies cited, nonprofessional personnel were quickly taught to administer the program. Another important feature of the procedures was the use of extrinsic reinforcers in the form of tokens which could be exchanged for money. The money, in turn, could be used to purchase goal objects of S's choice.

The present study was designed to further validate the Staats' procedures and to demonstrate their generality by extending their application to a population of pre- and early-adolescent retarded readers, including 16 experimental and 16 control Ss from regular fourth-, fifth-, and sixth-grade classrooms, and 8 experimental and 8 control Ss of equivalent chronological age from special classes for the mentally retarded. The need for monetary reinforcement in conjunction with the token reinforcer system was also tested. The program was administered by housewives recruited from the community.

In general, the 24 experimental Ss participated in the study with excellent cooperation, attention, and work behaviors, thus substantiating the motivational effects of the extrinsic reinforcers. The mean number of reading responses made during the experiment was over 51,000. A mean of 550 new (unknown) words was presented with an average retention rate of 79% of these words on a short-term retention test and 54% on a long-term test.

The results showed that Ss taught by the Staats procedures, using nonprofessional administrators, scored as high on the Lorge-Thorndike Group Intelligence Test and the California Reading Test as did the control Ss. Also, experimental Ss reinforced with tokens performed as well as experimental Ss reinforced with tokens and money. The experimental Ss scored significantly higher on a 100-item word recognition task, and made a smaller percentage of errors, in a prose reading task than did control Ss. There were no effects of type of reinforcement upon performance in the word recognition or the prose reading task.

On those measures internal to the procedures (such as error rate, number of responses, etc.), the performance of experimental Ss from regular classrooms receiving monetary reinforcement again did not significantly differ from the performance of the corresponding experimental Ss receiving token reinforcement. Comparisons between the two experimental groups of educable mentally retarded Ss did show differences in favor of those Ss receiving monetary reinforcement on measures of response rate.

It was concluded that the procedures are generally amenable to use by these types of Ss and to administration by nonprofessional personnel. The addition of monetary reinforcement to the token reinforcer system did not have significant effect on the performance of Ss from regular classrooms, but appeared to have an effect on Ss from classes for the mentally retarded.

I
INTRODUCTION

The present study is part of a continuing program of research directed by A. W. Staats. In conjunction with this research program Staats has suggested a general learning theory of human behavior called an integrated–functional learning approach (Staats, 1965a, 1965b, 1968; Staats & Staats, 1963). One feature of this approach involves an integration of the basic principles of classical and operant conditioning and their elaboration into complex theoretical mechanisms capable of describing and explaining complex behaviors.

A second feature of an integrated–functional learning approach is the concern for choosing samples of behavior for scientific study which are more representative of the actual behaviors of greatest importance. In conjunction with these considerations the second author and his associates have engaged in a variety of research projects. The goal of one set of studies was to answer several questions concerning some of the basic mechanisms of language learning and to delimit some of the stimulus functions fulfilled by words (cf., Staats, 1955, 1957a, 1957b, 1961). A second set of studies involved analyzing reading in behavioristic terms and devising laboratory procedures which would allow the study of a complex problem such as reading acquisition (Staats, 1965a, 1968; Staats & Butterfield, 1965; Staats, Finley, Minke, & Wolf, 1964; Staats, Minke, Finley, Wolf, & Brooks, 1964; Staats & Staats, 1963; Staats, Staats, Schutz, & Wolf, 1962). Certainly, reading acquisition and reading–problem remediation are two members of the class of behaviors which one would hope to be able to explain on the basis of sound psychological principles.

On the basis of previous findings and analyses, a method was devised for the treatment of children who, because of learning problems, have not been able to read normally. The method, which involved a system of extrinsic reinforcement employing tokens, was initially tested in a study using an adolescent male identified by the experimenters as a culturally deprived juvenile delinquent (Staats & Butterfield, 1965). The subject participated for 40 hours of training over a 4 1/2–month period. During this time he worked and attended well in the experimental training program, making over 65,000 word reading responses. He received special training on 761 words he did not know, retaining 430 or 57% of these words on a retention test. He passed all of his courses in school (the first time he had ever passed a course in his whole history); his misbehaviors in school fell off markedly; and his general attitudes toward school and school work appeared to improve. In addition, achievement tests showed an increase from the 2.0 grade level in reading to the 4.3 level.

These results suggest that the reinforcer system maintained the subject's behavior over an extended period of time. This is important when studying the acquisition of behaviors which require considerable time to evolve.

A second feature of the above study is that the procedures used were sufficiently simple and explicit that they could be administered by a nonprofessional. This is important in the light of the high cost of running a program utilizing a one-trainer-to-one-student ratio when professional personnel are the trainers.

The amenability of the procedures to administration by nonprofessional personnel and the strength of the reinforcer system received a further test in a study by Staats, Minke, Goodwin, and Landeen (1967). In this study 18 experimental subjects from regular classrooms and special classes for the mentally retarded (including some subjects judged by the experimenters to be mentally disturbed) at the junior high school level were presented with the same materials and procedures used by Staats and Butterfield (1965). In this study, the type of instructional technician was varied with half the subjects receiving the reading instruction from adult volunteers and the other half of the

subjects from high school students. The subjects participated for an average of 38.2 hours of training.

In the initial plan of this study, the 18 experimental subjects were paired with 18 control subjects who participated in the pre- and posttests but did not receive the remedial training. However, about half way through the experiment the schools involved started special reading programs supported by the Elementary and Secondary Education Act Titie I. Since several of the control subjects participated in this program, the control group data were contaminated, making the interpretation of the standardized test scores difficult. The experimental group did not differ significantly from the control group on the standardized tests used. Along with the contamination of the control group data, other reasons were given which might possibly account for the lack of differences between the groups on these measures. These included the following suggestions: (1) the tests used were not sufficiently sensitive to measure those changes which had taken place because of the program; (2) the changes brought about by the program did not transfer to the test situation; and (3) the test-taking behavior of the subjects was inappropriate and consisted of guessing and random marking of the answer sheets. The authors favored the third alternative. On a 100-item word recognition test the experimental group did show significant gains over the control group.

An examination of the data from the experimental group derived from their performance on the remedial reading materials showed that they made an average of 94,425 word reading responses. The 18 subjects were presented with an average of 593.5 words they did not know. Of these words, on which the subjects received special training, 469.7 (about 86%) were retained on a short-term retention test, and 343.5 (about 68%) were retained on a long-term retention test. The performance of the group as a whole surpassed the performance of the single subject in the Staats and Butterfield (1965) study on these criteria.

It was observed that as the program progressed the subjects' reading rate increased, even though the materials increased in difficulty. Also, the amount of reinforcement per response decreased over time. The attention, attendance, and cooperation of the subjects were maintained during the training period.

A third study using these remedial reading procedures was conducted by Judith McBurney (1967). This study was designed to extend the procedures to a population of pre-adolescent, emotionally disturbed males. Three experimental and three control subjects were selected. Extensive testing was done to ascertain the changes in academic and psychological (adjustment) performance which occurred during the time the experiment was being conducted. The experimental subjects participated for 55 sessions of approximately 30 minutes each.

The results of the academic tests used showed a slight advantage in gain scores for the experimental group. Since the subjects were tested individually these test scores may be more reliable than the test scores in the Staats et al. (1967) study. The small number of subjects and the noting of various ceiling and floor effects in the tests used suggest the need for caution in interpreting these results. The differences between the experimental subjects and control subjects on the tests of psychological (adjustment) performance showed no clear patterns. On some tests the differences favored the control group and on others the experimental group was favored.

The performance of two of the three experimental subjects in the experimental procedures indicated that the reinforcer system was maintaining their behavior adequately with respect to attention, attendance, and cooperation. This finding is important when considered in connection with the criteria for identifying the emotionally disturbed which characteristically include intractability in regular classroom settings. The third experimental subject performed in an inconsistent manner from session to session. His response rate decreased as the experiment progressed. This is the only contradiction to the trend noticed in all other subjects who have so far been administered the procedures.

From the three studies reviewed to this point several important generalizations emerge. First, there is evidence that the overall reinforcer system used is effective with a variety of types of subjects. In the study of complex behaviors which require considerable time to acquire, it is important to devise a motivational system which will ensure adequate long-term involvement on the part of the subjects. However, the need for monetary reinforcement in conjunction with the other reinforcers inherent in the remedial reading procedures has not been tested. A second generalization is that the remedial reading materials and procedures appear applicable to a wide variety of types of subjects. A third feature of the procedures and materials is their amenability to administration by nonprofessional personnel.

The procedures of the remedial reading training involve several types of tasks. The details of the various phases of training will be described in the method section of this report. However, for purposes of comparison with other remedial reading programs they will be briefly outlined

here. The first phase of the training involves the presentation of all those words occurring in a given lesson which have not been previously presented to the subject as part of training. The words are presented singly. The second phase of the program presents these same words with other words already presented in earlier lessons; in this phase the new words are imbedded in complete paragraphs. In the third phase of the program the subjects read silently the whole story on which the lesson is based. The fourth phase is a comprehension phase in which subjects answer questions on the materials they have just read. Those words the subject did not know in the first phase are also presented in a later, retention task to ascertain how effective the single word training was.

A review of several other remedial reading training programs will help the reader to relate this study to the significant work already done in the area of remedial reading. This is not intended to divert the attention of the reader away from the main purpose of this study which is to demonstrate the feasibility of applying basic learning principles to the analysis and control of complex human behavior. The remedial reading programs described below are those believed by Staats to characterize the major orientations in the area.

Grace Fernald (Fernald, 1943; Fernald & Keller, 1921) suggested that the effect of providing auditory and kinesthetic cues as well as visual cues in letter or word recognition tasks would help retarded readers to observe words in a systematic and accurate manner. These extravisual cues were provided by having subjects write, trace and pronounce letters and words. She contended that by providing this whole constellation of interrelated cues for each troublesome letter or word, discrimination would be aided. A program was devised to implement the use of these extravisual cues.

The first step of this program involves the presentation of single words with extensive visual, manual, and auditory practice. Next, these words are presented in phrases with visual, manual, and auditory practice. The next three steps involve progressing from reading simple stories made up from the phrases practiced in Step 2, through reading paragraphs silently for comprehension.

Gates (1949) suggested yet another set of cues to aid the reader in correctly identifying words. Although several types of exercises were outlined, in general, the exercises may be described as providing contextual cues "...the setting for each new word should be carefully worked out so that the word is surrounded with such abundant and suggestive context cues that the pupil will be quite sure to figure out [the correct word] promptly and correctly [Gates, 1935, p. 272]." The exercises include the use of other words to provide meaning cues, pictures, coloring, and cutting and matching work.

Gates also stated that as a general theory there is no distinction between regular developmental classroom teaching and remedial teaching. The difference in practice is one of degree with individual differences of more importance when instructing a retarded reader. Other authors have ascribed to the same point (cf., Otto & McMenemy, 1966; Witty, Freeland, & Grotberg, 1966).

The teaching methods of Monroe (1932) differ sharply from the two general methods previously reviewed. Her emphasis was upon the specific treatment of particular difficulties such as the addition or omission of sounds; i.e., phonetics form the basis of the approach.

The initial training in the Monroe procedures involves the pairing of letters and their most frequent sound. As part of the training procedures the subject traces the letter while making its sound. The next step is to combine consonants and vowels into simple, regular words and then to develop the remedial work on the basis of the recognition of words from their sound components. The last step in the program is the reading of graded phonic stories based on the interests of the subjects.

Kirk and his associates (Kirk, 1940; Hegge, Kirk, & Kirk, 1945) have also used a phonetic approach to remedial reading with extensive drill in associating visual, manual, and articulatory cues with sounds.

A third type of phonic approach is exemplified by the work of those who have developed various phonetic alphabets. For example, the Pitman alphabet of 40 phonetic letters has been used to teach initial reading and, to some extent, remedial reading. Harrison (1964) describes several of these studies.

A fourth group of remedial reading programs includes the more general or eclectic approaches of such authors as Cole (1938), Durrell (1940), Schonell (1948) and Harris (1956). These approaches offer a variety of specific training techniques, motivational considerations, and material.

A more extensive review of the approaches briefly described above would show that many of the procedures and/or considerations mentioned by the various authors are concerned with the problem of motivation. The suggestion that the motivational procedures used in the two studies by Staats and his associates may be generally applicable is important in the light of the apparent need for some motivational

system which will support the remedial student in his efforts.

A second observation about the several remedial reading approaches reviewed of interest in conjunction with this study is that the materials and procedures of the programs progress through stages similar to those outlined in the methods section of this report. This is true even though the various programs were developed in a relatively independent fashion. The progression from (1) word recognition training (sometimes preceded by phonetic training) to (2) the recognition of words in context and on through (3) silent reading to (4) reading for comprehension is common to almost all of the programs reviewed. The specific procedures and materials of the various programs naturally differ, but often such a variety of materials and procedures are offered that there is still considerable overlap across programs. It is recognized that several of the programs add intermediate steps to the above outline, but in general they are built upon this basic framework.

Indeed, it is a basic tenet of behavioral engineering that if one examines a situation in which the behavior of interest is already functional, the present situation is likely to have many similarities to the theoretically ideal situation as defined by the engineer after analysis. To illustrate this suggestion, a study by Hewett, Mayhew, and Rabb (1967) will be reviewed. An experimental reading program for neurologically impaired, mentally retarded, and emotionally disturbed children was prepared by the Neuropsychiatric Institute School, University of California, Los Angeles. The program was an extension of a program developed by Bijou and his associates (Bijou, 1957, 1965; Bijou & Sturges, 1959; Birnbrauer, Bijou, & Wolf, 1963).

The underlying principles upon which the program was built included those of operant conditioning and programed instruction. The materials used were derived from the Ginn and Company Basic Readers. The procedures included (1) four types of single word training on the new words (words not yet learned in the program), (2) phrase reading (with the new words occuring amid words already presented in the program), (3) oral reading of the chapter in the Basic Reader which contained the new words just learned, (4) a set of comprehension questions, and (5) discrimination exercises with the newly learned words. Thus, with the exception of Step 5 (which may be considered as belonging to the general area of Step 1) the procedures are similar to the basic outline of the several remedial reading programs reviewed above.

A total of 26 subjects participated in the program. Of these, four were labeled as educable mentally retarded (EMR) and one was a preschool normal female. The remainder of the subjects were classified as neurologically impaired, autistic, or emotionally disturbed. The subjects received training for 25-35 minutes three times a week. If the performance level of a subject on the materials fell below 75% correct responses, he repeated the materials until that level was reached. If this criterion was not met after several repetitions, the subject was presented with an errorless training sequence which utilized very simple discrimination materials.

The reinforcer system involved extrinsic rewards such as candy, money, and toys. In the first three sessions one piece of candy (or in the case of two of the EMR subjects, one penny) was given for each correct response. For the next three sessions all subjects were put on a 1:5 reinforcement schedule. Later, the subjects went to a 1:10 schedule with the exchange delayed until the end of the sessions. A system in which checkmarks were given were used to keep track of the number of correct responses. The checkmarks were given on a 1:1 basis. Eventually, a 1:200 schedule for tangible rewards was established with the subject being given his choice of one 5-cent prize for each full card of 200 checkmarks.

The subjects were moved through the different reinforcement schedules at individual rates. The judgment of a trained psychologist determined when a subject was presented a new reinforcement schedule. The authors concluded that the reinforcer system maintained adequate attention, rate of responding, and sustained progress with all subjects. The subjects participated for varying numbers of sessions because of differences in the length of hospitalization and the availability of out-patients. The range of participation was from 8 to 137 sessions.

As part of the overall research program the authors explored the effect of reinforcement variables. The three variables investigated were the removal of candy, taking away candy for errors, and instituting a "time-hold" (the subject was given a set number of seconds in which to respond and no response was considered an error).

It was concluded that the removal of the candy reinforcer had no effect. "Candy or no-candy did not appear a significant factor in determining this boy's functioning, and observed changes probably were due to other factors. Similar results (candy not a significant variable in determining subject's performance) had been obtained with several other children subjected to the same conditions [Hewett et al., 1967, p. 46]."

Taking candy away from subjects when errors were made did not produce much difference compared with the withholding of rewards for incorrect responses. The "time-hold" variable did produce an increase in the rate of responding, but not in the accuracy of response.

On the basis of previous findings the following hypotheses were formulated for test in this study:

Since the procedures used in the present study do not differ too greatly at an abstract level from the sequence of steps used in other already existing remedial reading programs (although in this case the procedures are amenable to administration by nonprofessional personnel), it was hypothesized that there would be no significant differences, as measured on standardized achievement tests, between the experimental groups using the Staats' procedures and the control groups who would be receiving regular developmental reading training and/or special attention from the regular classroom personnel. Further, it was hypothesized that differences would be found in favor of the experimental groups on nonstandardized measures of high face validity.

It was also hypothesized that there would be no differences in test performance between the groups receiving monetary reinforcement in conjunction with the token reinforcer system and those groups receiving only nonmonetary reinforcement (points, etc.).

Regarding experimental groups only, on those measures internal to the procedures (such as error rates, number of responses, etc.), it was hypothesized that there would be no differences between the performance of the subjects receiving monetary reinforcement and the corresponding subjects receiving nonmonetary reinforcement.

II

METHOD AND PROCEDURES

SELECTION AND ASSIGNMENT OF SUBJECTS

Twenty-four experimental and 24 control subjects were selected to participate in the experiment. Of the total 48 subjects, 16 were from special classes for educable mentally retarded students (EMR Ss) and 32 were from regular classrooms (NRR Ss). The selection of subjects proceded in the following manner: The building principals of the schools involved in the experiment compiled a list of those students in the fourth, fifth, and sixth grades or, in the case of the EMR Ss, students of equivalent chronological age who were identified by their teachers as deficient in reading skills. All on these lists were tested using a 100-item word-recognition task (100-WRT). (For more information about this task see the section on Instruments.) Those students with scores between 9 and 70 were further tested using the California Reading Test, Upper Primary Level, Form X, and the Lorge-Thorndike Intelligence Test, Level 3, Form A, Verbal Battery. The scores of 9 and 70 were used as a lower and upper limit, respectively, because earlier research has shown these limits to be necessary. Subjects with scores below 9 on the 100-WRT tend to have an unusually difficult time with the reading materials. Likewise, Ss with a score above 70 on the 100-WRT tend to have an unusually easy time with reading materials. The main variable used for selection was the 100-WRT Test scores.

Table 1 shows the manner in which the Ss were distributed across schools. It should be noted that with a limited number of participating schools and appropriate Ss the assignment to experimental and control groups was not strictly random, nor was it possible to match Ss in any completely satisfactory manner and then assign the Ss within pairs in a random manner.

Subjects were assigned to groups according to the following considerations: (1) try to make the groups as comparable as possible on the pretest scores with the 100-WRT Test score being considered most important; (2) try not to assign Ss in the same school to two or more conditions which might interact in an uncontrolled and unpredictable manner (for example, to have some Ss in a school receive material reinforcement and others who do not for the same performance would, a priori, appear likely to cause severe interactive effects); and (3) keep in mind the problems of logistics and the availability of instructional technicians in the school area.

The Ss were assigned to eight different treatment groups. The eight groups were defined by all possible combinations of three dichotomous factors — type of subject (retarded readers from regular classrooms or from special classes for the mentally retarded), type of reinforcement (experimental Ss earned tokens which were redeemable for money by one-half the experimental Ss and for points by the other half; see the following section for a description of the reinforcement procedures with the control group Ss) and treatment (experimental Ss received the Staats' procedures while control Ss received instruction from the regular school personnel).

In Table 2 are shown the group means and standard deviations of the experimental and control groups with respect to the pretest scores on the 100-WRT Test, the Lorge-Thorndike Group Intelligence Test and the California Reading Test. There was considerable variance among subjects and groups. An inspection of the group means shown in Table 2 reveals the extent to which it was possible to minimize the differences between groups. Analysis of covariance techniques were used to analyze the posttest scores using the corresponding pretest scores as the covariates.

SPECIAL PROBLEMS WITH SUBJECTS

Some of the control Ss were paid for their "participation" in the experiment. These Ss were

Table 1

Breakdown of Ss by Groups Within Schools

Schools	EXPERIMENTAL Ss				CONTROL Ss				TOTAL
	EMR		NRR		EMR		NRR		
	$[a]	Pts[b]	$	Pts	$	Pts	$	Pts	
School I			8					7	15
School II	4			3		4	2		13
School III		4			4				8
School IV				5				1	6
School V							6		6
Totals	4	4	8	8	4	4	8	8	48

[a]Denotes the Ss receiving monetary reinforcement.

[b]Denotes the Ss receiving nonmonetary reinforcement.

Table 2

Means and Standard Deviations of Pretest Scores by Variable and Groups

Groups	SRA-100 Test (100-WRT)	Lorge-Thorndike (DIQ)	California Reading Test Total Reading
EMR-$-Exp[a]			
M	21.250	65.000	2.525
SD	8.098	8.367	0.386
EMR-$-Con			
M	35.250	68.500	2.950
SD	28.547	4.041	0.656
EMR-Pts-Exp			
M	15.000	63.500	2.425
SD	7.394	9.983	0.250
EMR-Pts-Con			
M	32.750	71.500	2.850
SD	20.189	7.047	0.507
NRR-$-Exp			
M	62.250	78.125	3.537
SD	9.982	11.594	0.288
NRR-$-Con			
M	58.286	86.429	3.714
SD	10.935	6.997	0.385
NRR-Pts-Exp			
M	56.250	81.125	3.387
SD	10.068	13.410	0.344
NRR-Pts-Con			
M	56.625	79.500	3.600
SD	18.260	10.876	0.689

[a]Denotes the group of Ss who were from the EMR classes who received monetary reinforcement ($) and who received the Staats' procedures (Exp).

labeled as "paid-control" Ss. Each paid-control S was paired with an S from the experimental group which received material reinforcement. The paid-control Ss were told at the start of the experiment,

> You are participating in a reading experiment conducted by the University. We are interested in how well you read. For the next several months I will visit you and talk with you about your reading activity. When I come I will have varying amounts of money to pay you for your reading and participation in this experiment. Remember you are to read as much and as well as you can, both at school and at home.

Whenever the experimental S earned an amount sufficient to purchase the object for which he was working, the same amount was "earned" by the corresponding control S. There was a time lapse of three weeks between the earning of an object by the experimental S and the paying of the control S to insure the same interval between pay-offs for experimental and control Ss and to allow accurate bookkeeping.

Whenever a paid-control S was to be paid, the author went to the school where he was enrolled and arranged to have him called out of class. The S was reminded of the reason for which he was receiving the money by having E paraphrase the instructions given initially (see above). A receipt was signed by S. The S was instructed to take the envelope with the money to the school secretary to leave it there until school was out. This was to prevent S's playing with the money in class and to minimize the chance of losing it.

The Ss labeled "control-points" in this experiment participated only in the pre- and posttesting. Otherwise, they had no contact with E.

One NRR paid-control S was lost from the experiment in the seventh week because she was severely injured in an accident and did not return to school. One S in the EMR-Experimental-Paid group was lost after he had completed 57 of the 65 desired sessions. He was discontinued because the family moved from the state. His mother did bring him back to Madison so that he could participate in the posttesting.

INSTRUMENTS

The primary test used, both for selection and as a dependent variable, consisted of 100 words selected randomly from the reading materials used in the experimental sessions. This test was called the 100-item word recognition test (100-WRT). Twenty words were selected from every other level of difficulty presented in the materials (1.2, 1.7, 2.3, 3.0, and 4.0 levels) out of a pool of 4253 total words presented in the program. The words were typed on 3 x 5 index cards. The 100-WRT Test was administered individually to each S. Subjects had to correctly pronounce the word shown them on the card in order to receive credit. The order in which the words were presented was arranged such that every fifth word was from the same grade level. Thus, the first, sixth, eleventh, etc. words were from the 1.2 grade level; the second, seventh, twelfth, etc. words were from the 1.7 grade level; and so on.

The 100-WRT Test was initially administered before any of the Ss were assigned to groups. The same test was administered at the termination of the experiment.

Further pretests given included the Lorge-Thorndike Intelligence Test, Level 3, Form A, Verbal Battery, and the California Reading Test, Upper Primary Level, Form X. The alternate forms (B and W, respectively) were given as posttests.

A third type of test situation used in the analysis of the effects of the program involved having each S read a story from the SRA Reading Laboratories, Lab 1c, Olive (The 5-Minute Story Test). The stories were printed on 8 1/2 x 11 paper. The number of words read per unit time and the percentage of mistakes constituted the data. The experimental Ss read one set of stories at the beginning of the experiment and another set at the end. It was not possible to get a pretest measure on all of the control Ss. Therefore, for them the test was given at the end of the experiment only. The stories used as a posttest were exactly the same for the control and experimental S.

INSTRUCTIONAL TECHNICIANS

The instructional technicians (the adults who administered the Staats' procedures to the Ss) were selected by having the building principals of the participating schools provide the names of adults living in their school district who had demonstrated a willingness to aid the school in various projects and who had shown themselves to be generally responsible. Of the 64 adults initially contacted, 24 were interested enough to attend a meeting. Those women interested in participating in the experiments attended a second training meeting in which they had the chance to administer the program under the direction of the experimenter. In this meeting the instructional technicians were further instructed in the manner of keeping data. Each potential instructional technician attended the second series of meetings for about 1 1/2 hours.

A third series of meetings was held for those potential instructional technicians who felt the need for more training. Only six women attended these meetings.

Those instructional technicians living in the same school districts as Ss were assigned at random to the Ss in their district. However, because of an unequal distribution of Ss and instructional technicians across school districts it was not possible to assign all of the instructional technicians to Ss in their school districts. Thus, some instructional technicians traveled several miles each day to administer the program to the Ss. Because of the time and inconvenience involved in going so far, as well as the cost in transportation and babysitter expenses, some of the instructional technicians administered the program to two Ss each day. Care was taken to ensure that this occurred equally often across the experimental groups. The instructional technicians were paid $2.00 an hour to cover their costs.

The average age of the instructional technicians was approximately 42 years. They had completed an average of 13.3 years of school. The mean number of children in their household was 2.85, and they were the mothers of an average of 3.4 children. Their husbands had completed an average of 13.6 years of schooling. The mean family income for the preceding year was $9,500. On the basis of these descriptive statistics it appeared that the average instructional technician was a mature woman who cared for an average-sized family. Both she and her husband had an above average education. The family income was also above average. (All comparisons were made using data from the Statistical Abstract of the United States, National Data Book and Guide to Success, U. S. Dept. of Commerce, 87th edition, 1966).

The instructional technicians were frequently observed throughout the course of the study. A special attempt was made for the experimenter to be present at the first few sessions when the instructional technician first started working with her S.

DESCRIPTION OF INSTRUCTIONAL MATERIALS

The materials used in this experiment were taken from the SRA Reading Laboratory materials, Labs Ia, Ib, Ic, and IIa. The SRA Reading Laboratory consists of stories developed for and grouped into grade levels. For the purpose of this study, stories were taken from the 1.2 (20 stories), 1.4 (60 stories), 1.7 (20 stories), 2.0 (32 stories), 2.3 (12 stories), 2.6 (12 stories), 3.0 (32 stories), 3.5 (32 stories), and 4.0 (10 stories) grade levels. Once a particular Reading Laboratory was selected for inclusion at a given grade level, all the stories at the grade level were presented in sequential order (with the exception of the 4.0 grade level, where only the first 10 stories from Laboratory IIa were presented). The different number of stories at the various levels was a function of the attempt to control the rate of introduction of new words.

It should be mentioned that materials other than the SRA Reading Laboratories could have been used. However, the SRA stories have several features which make them attractive for the purposes of this study. First, there is a somewhat controlled introduction of new words in each lesson. Second, a set of comprehension questions is available for each story. In addition, as grade level increases, the stories become more difficult, they are longer overall, and paragraph length increases. This resulted in a gradual decrease of the average amount of reward delivered for each response.

The first step in adapting materials for this reading training program was to make a running list of new words that appeared in the stories. Each word, or form of a word (past tense, plural, etc.), that had not occurred in a previous story was added to the list. From this list the new words that occurred in a given lesson were typed on a separate 3 x 5 index card. Each paragraph was typed on a 5 x 8 index card. The complete story and the comprehension questions were typed on separate 8 1/2 x 11 sheets of paper. These comprised the materials which were presented to the student during the training. In addition, data sheets were prepared for each lesson. These data sheets were used by the instructional technicians.

In addition to the individual lessons, each instructional technician was furnished with a packet of materials which included pencils, paper clips, rubber bands, Vocabulary Review data sheets, 100 each of yellow, red, and blue tokens, and charts for plotting the money or point value of the tokens delivered each day. The uses to which these materials were put are illustrated and discussed at the appropriate time in the following sections.

PROCEDURES

Each lesson consisted of four phases: Individual Word Phase, Oral Reading Phase, Silent Reading Phase, and Comprehension Phase. These phases are individually discussed in detail below.

Individual Word Phase

The first phase consisted of the individual presentation of each word that occurred in the lesson being covered and that had never before

been presented in the program. Each word was presented on a separate 3 x 5 index card. The words were presented to S one at a time, and S was asked to read them.

The first time through the stack of word cards for a given lesson, every correct reading of a word was rewarded with a yellow token (worth 1/5 of a cent or two points). Whenever a card was read correctly, it was removed from the stack. If S read a word incorrectly, or did not read it at all, the instructional technician told S what word was typed on the card. The S was then to repeat the word while looking at it. If S was not looking at the word when he responded, he was required to look at the word and say it again. When S had properly responded to the card after being prompted, it was returned to the group of cards still to be presented. No token was delivered at that time.

After having gone through the original stack once using this procedure, the stack (with the correctly given words removed) was presented again to S, and the procedure was repeated (i.e., words which S read incorrectly were prompted, but not rewarded, and then were returned to the stack for later presentation). On the second, or subsequent presentations of the words, a blue token (worth 1/10 of a cent or one point) was awarded contingent upon a correct reading response. When there were no more cards in the stack, i.e., when S had read each word correctly once, without prompting, the Individual Word Phase was ended and the Oral Reading Phase was instituted.

For information concerning the data-marking procedures for this phase and the other three phases the reader is referred to the administration manual (Staats, Van Mondfrans, & Minke, 1967).

Oral Reading Phase

Each paragraph of the story being presented was typed on a 5 x 8 index card. These paragraphs were presented in the order they appeared in the story. When S read the entire paragraph without an error, he received a red token (worth 1/2 cent or 5 points). When S read a word incorrectly on the card, the instructional technician pointed to the word, and S was to repeat the word while looking at it. After S finished a card upon which errors occurred, the card was put aside (with no token delivered) and presented again after the remainder of the cards had been presented. When errors occurred on two or more cards the first time through, all the missed cards were presented again before a third trial on any card took place. A paragraph was repeated until S had read it through in its entirety without an error, at which time a yellow token was earned. When all of the paragraphs in a story had been read correctly, the tokens which S had earned were given and the next phase of the training was initiated.

Silent Reading Phase

Following the Oral Reading Phase, S was given a sheet containing the story appropriate to the particular lesson upon which he was working. These stories were typed on regular 8 1/2 x 11 sheets of paper. The S was instructed to read the story silently and was told that it was important to read to understand the story so that he could answer questions.

The instructional technician watched S closely to insure that he was actually reading the story. If it was apparent that S was not carefully attending to the story, the instructional technician required him to read aloud during this phase on subsequent lessons. It was found that reading aloud in this phase for several sessions resulted in adequate atiention on the part of S so that he subsequently was allowed to read silently.

The S was given four yellow tokens upon the completion of the story, at which time the next phase was instituted.

Comprehension Phase

The comprehension questions used were those which accompanied the SRA stories adapted for this program. They took one of two forms: Some of the questions required a written answer, and some were of the multiple-choice type. In either case, S wrote his answers (or circled his answers) to the comprehension questions and then handed them to the instructional technician. For each correct answer, a red token was earned. If a spelling error had been made when a written answer was required, S was to correct his answer, and then a yellow token was earned. For incorrect answers S reread the appropriate paragraph and corrected his answer, at which time a blue token was earned. All the tokens earned were presented at the end of the phase. The end of this phase marked the completion of the lesson. At this point the whole procedure was repeated with the next lesson in the series.

Vocabulary Review

Some of the words presented in the Individual Word Phase of the procedure were words that S already could read. Many others, however, were words that the procedure was intended to teach. Short-term retention of these new words was assessed through an analysis of errors in the Oral Reading Phase. The Vocabulary Review assessed the long-term retention of words which were initially unfamiliar to S. In most cases, the Vocabulary Review was presented following every twenty reading lessons. However, with

some Ss, particularly the EMR Ss, so many words were missed that the Vocabulary Review was given every ten lessons. The Ss who necessitated the change were the slower Ss, so the actual span of time between Vocabulary Reviews for them was similar to the span of time for the faster students who received the Vocabulary Reviews after every twenty lessons. The Vocabulary Review dealt only with the words which were presented in the immediately preceding twenty lessons (or ten lessons) and not in all the lessons presented to that time.

The materials necessary for the Vocabulary Review were prepared from the materials used in the Individual Word Phase of each lesson. Each day, after S had left, the instructional technician set aside the individual word cards on which errors had been made during that session. These words were itemized on the Vocabulary Review data sheet. By doing this daily, the data sheet was completed by the time the Vocabulary Review was presented.

At the end of each twenty lessons (or ten lessons) the instructional technician combined the cards which had been set aside from each lesson into a single stack (in the order in which the words appeared on the Vocabulary Review data sheet). The Vocabulary Review was then administered in the same manner as the Individual Word Phase of the individual lessons.

The Token System

An important aspect of this study was the use of a motivational system consisting of the awarding of tokens, contingent upon correct responding, which were exchangeable for various rewards. Three colors of tokens were used, each color representing a different value. Thus, a blue token was worth 1/10 of a cent or 1 point; a yellow token was worth 1/5 of a cent or 2 points; and a red token was worth 1/2 of a cent or 5 points.

Each experimental S in the material reinforcement groups, indicated to his instructional technician some object for which he would like to work, such as a phonograph record, hair cream, a pair of shoes, and so on, and the item's exact price. When S had earned tokens whose monetary value equaled the amount necessary for the purchase of the item, he was given the money and allowed to buy the item.

At the end of each session the monetary or point value of the tokens earned was computed using computational guides provided on a graph and added to the amount which had been earned in previous sessions. This amount was then plotted on the graph. For best visual effect, a bar graph was used. For the experimental groups receiving monetary reinforcement, each time S earned his payoff a new progress chart was started. If, on the day the S reached his criterion, he earned tokens worth more than the object for which he was working cost, the additional amount was entered immediately on the new chart. For the experimental groups receiving only points, the point value of the tokens was entered on the graph. This constituted the only formalized reinforcement other than tokens these Ss received.

The proper use of the graph involved entering S's progress in his presence, with substantial positive comment and approval on the part of the instructional technician. Regardless of the amount of money or points S had earned during the session, he was made to feel that he had performed well.

Bonus System

In the earlier sessions the stories were relatively easy and Ss could earn many tokens. As an S progressed in the program, the stories became harder and longer, and more responses were required per unit reinforcement in the Oral Reading, Silent Reading, and Comprehension Phases. In order to lessen the abruptness of the drop in earnings a bonus system was introduced. Those instructional technicians whose Ss were earning considerably less than 200 points or 20 cents per day were instructed to reward any sequence of unusually good behavior with an extra token. For example, if an S consecutively read several difficult words in the Individual Word Phase without an error the instructional technician would give an extra red token. These bonus tokens were tabulated separately by the author but were included in the graphs and totals by the instructional technicians. The total amount earned by the S would be the sum of the bonus tokens and the tokens actually earned by participating in the regular program.

The Session

Although the decision as to how long to work with Ss each day was a somewhat arbitrary one, for various reasons a one-half hour period of actual reading each day was the length of the session for this research. This one-half hour period corresponded conveniently with the periods in school, it is the time period used by earlier researchers in this program (Staats & Butterfield, 1965; Staats et al., 1967), and this time period provided the instructional technicians with sufficient time to set up and clean up materials, etc. and still not spend more than an hour away from home. This one-half hour period included only the time spent on the actual reading of the lessons and did not include the time required to set up the materials, enter the day's earnings on the graph, and so on.

Usually, more than one lesson was covered in any one session. Occasionally, S had progressed only part way through a lesson before the time elapsed. There were two permissible stopping places within a lesson—after the Individual Word Phase or after the Oral Reading Phase. Stopping after the Silent Reading Phase would result in an inflated error rate in the Comprehension Phase. When a lesson was split across two sessions, two data sheets were used to facilitate the compilation of the data.

III
RESULTS

THE RESPONSE DATA INTERNAL TO THE PROGRAM

The reading responses of each S in the experimental groups were carefully tabulated. The performance of the four experimental groups were compared on various measures such as total number of reading responses made, number of words missed the first time they were presented, and number of words missed after training.

Each experimental S participated in 65 training sessions of approximately 30 minutes duration (with the exception of one S in the EMR-Money-Experimental group who finished 57 sessions before he was terminated in the program because of moving).

The Total Number of Reading Responses

The total number of reading responses made by Ss varied greatly. The Ss in the EMR-Money-Experimental group made an average of 40,429 reading responses (SD = 3,971.74). This was 68% more reading responses than the mean number of responses made by the Ss in the EMR-Points-Experimental group ($\bar{X}$ = 24,045; SD = 12,791.86). The comparison of the two NRR groups showed that the NRR-Points-Experimental group made 9.3% more reading responses on the average ($\bar{X}$ = 62,779; SD = 14,815.96) than the NRR-Money-Experimental group ($\bar{X}$ = 57,418; SD = 10,875.75). The above differences were analyzed using covariate techniques with the 100-WRT pretest scores as the covariate (see Table 2) and the group means for the total number of reading responses made as the criterion variable. As can be seen in Table 3, the effect of Type of Subject was significant, $F(1,19) = 4.45$, $p < .05$. The interaction between Type of Subject and Type of Reinforcement was also significant, $F(1,19) = 8.418$, $p < .01$.

The cumulative group averages of the total number of reading responses made per session are presented in Figure 1. These curves include the words read in all the phases of the program. As can be seen from the figure, the curves for three of the four groups (the EMR Points group excepted) show positive acceleration (at least during the first 40 sessions). This is true even though as Ss progressed in the program the reading materials became more difficult. The curve for the EMR Points group is essentially a straight line.

Table 3

Analysis of Covariance Summary Table with Total Reading Responses as Criterion and 100-WRT Pretest Scores as the Covariate

Source	df	MS	F
Type of Subject (S)	1	281,624,747.375	4.447 *
Type of Reinforcement (Rf)	1	159,289,062.938	2.515
S x Rf	1	533,110,556.625	8.418 **
Within cells	19	63,326,816.303	

* $p < .05$
** $p < .01$

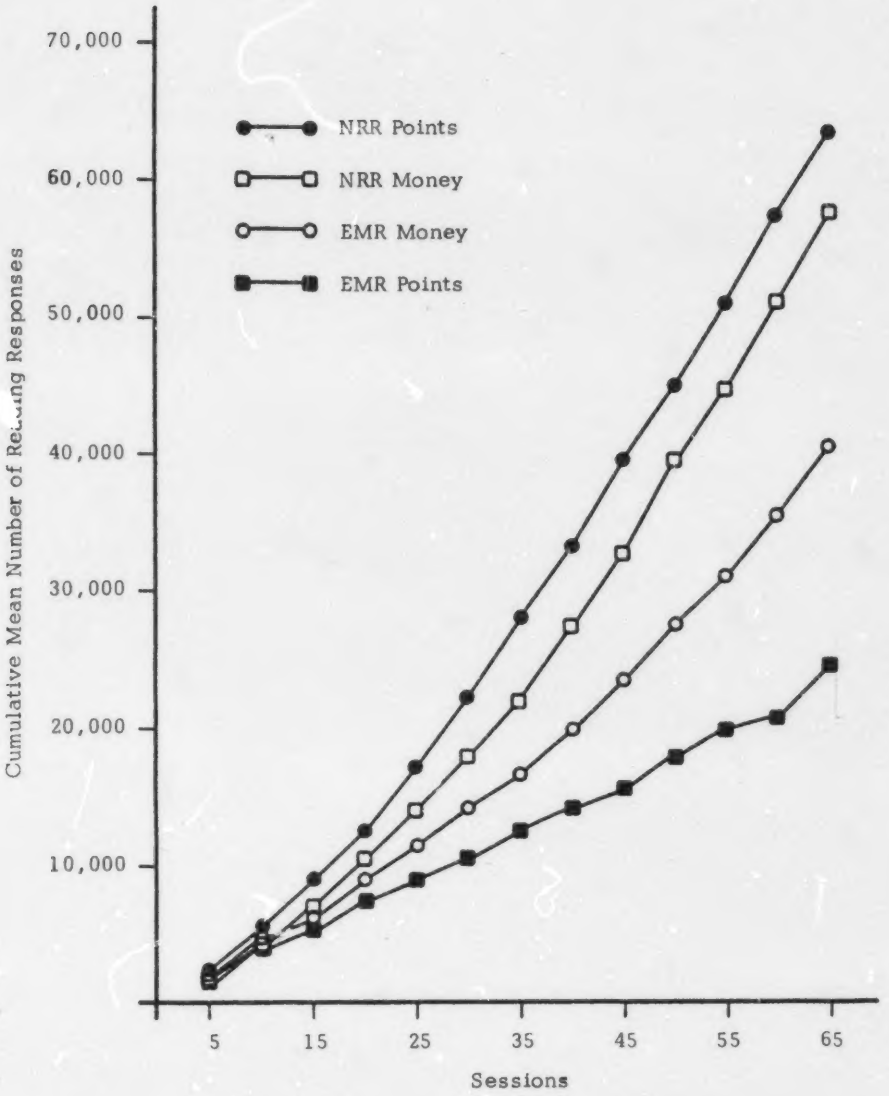

Figure 1. The Cumulative Mean Number of Reading Responses as a Function of Sessions Completed for the Four Experimental Groups.

Words Known upon First Presentation and on Retention Tests

Records were kept of the number of words the Ss missed on first presentation, the number of these words which were then later missed in the Oral Reading Phase, as well as the number of the words originally missed that S could not read when presented in the Vocabulary Review.

The Ss in the EMR-Points-Experimental group were presented with an average of 1001.25 different words in the Individual Word Phase during the course of this experiment. Of these, an average of 442.75 were known to the Ss (44%). The remainder of the words were missed initially, and the Ss were trained on them in the procedures. When these words were presented in the immediately following Oral Reading Phase an average of 72% of the words were retained by the Ss. On the long-term retention test (Vocabulary Reviews) the average percent retained was 40%.

The corresponding figures for the EMR-Money-Experimental group were higher in two of the three comparisons. These Ss were presented with an

average of 1490.5 different words of which 807 or 54% were initially known. On the first, short-term retention tests 69% of the words Ss were trained on were retained. The data from Vocabulary Reviews showed that 46% of the words initially missed in the Individual Word Phase were retained on this long-term retention test.

The Ss in the NRR-Points-Experimental group were presented with an average of 2119.6 words in the Individual Word Phase. They knew an average of 1462.1 (74%). The percentages retained on the two retention tests were 72% on the short-term test and 63% on the long-term test.

The Ss in the NRR-Money-Experimental group were presented with an average of 2102.1 words of which they knew an average of 1703.17 (81%) initially. On the short term retention test 83% of the words the Ss were trained on were retained. The percent retained on the long-term retention test was 67%.

The Amount of Reinforcement Per Reading Response

As the number of lessons the Ss had read increased, the length of the lessons increased. This was particularly true of the Oral Reading Phase and Silent Reading Phase of the lessons. The result of this increase was a gradual drop in the amount of reinforcement per reading response across sessions. The average number of points earned or the average amount of money earned in a session by a group was divided by the number of reading responses made in that session. The resultant ratio was called the "payrate." Since some Ss received money for the tokens awarded and other Ss received points, the payrate could be expressed in cents or points. In Figure 2 the average payrate for each experimental group across sessions is given. The numbers on the ordinate are in cents per response. For example the number .05 on the ordinate should be read "five-one hundredths of a cent," and would be equivalent to five-tenths of a point.

From Figure 2 it is clear that the payrate for three of the four experimental groups (EMR-Points group excepted) decreased as the experiment progressed. The curve for the EMR-Points group is quite irregular.

RESULTS OF TEST DATA

Because of the relatively gross differences between Ss within groups, the test data were analyzed using analysis of covariance techniques. Table 2 shows the group means for the pretests. In Table 4 the group means for the posttest scores and the gain scores are shown for all the tests given. The gain scores are defined as the mean score of a given group on a particular posttest minus the group mean on the corresponding pretest. The sign is retained to show the direction of the differences. The two scores analyzed from the Five-Minute Story, the percent of errors made (%-Errors) and number of words read per minute (W/Min) did not have a corresponding pretest for the control groups so no gain scores are given.

The Lorge-Thorndike Group Intelligence Test

An analysis of covariance with the posttest scores as the criterion and the pretest scores as the covariate for the Lorge-Thorndike DIQ scores, showed no significant differences (Table 5). The group means for this instrument are presented in Tables 2 and 4 for the pretest and posttest scores, respectively.

The California Reading Test

An analysis of covariance summary table for Total Reading scores from the California Reading test is presented in Table 6. The group means analyzed are presented in Tables 2 and 4. The three-way interaction involving the effects of Type of Subject, Type of Reinforcement, and Treatment approached significance, F (1,38) = 4.089, F = 4.10 needed at .05 level. This reflects the fact that the mean gains evidenced by the various experimental and control groups were not uniform in magnitude. The largest mean gain was made by the NRR-Experimental-Points group (.463 years). The smallest mean gain was made by the EMR-Control-Money group (.150 years). The second largest gain involved the NRR-Control-Money group with the second and third smallest gains involving the NRR-Control-Points and NRR-Experimental-Points groups, respectively. Thus, although the variance between the groups was significant, no consistent pattern was obvious. Comparisons of the experimental groups and the corresponding control groups showed that in two cases the difference in mean gains favored the experimental groups (EMR-Experimental-Money group over the EMR-Control-Money group by .175 years; NRR-Experimental-Points group over the NRR-Control-Points group by .276 years) and in two cases the control groups were favored (EMR-Control-Points group over EMR-Experimental-Points group by .100 years; NRR-Control-Money group over NRR-Experimental-Money group by .102 years).

100-WRT Test Results

In Table 7 the results of an analysis of covariance with the 100-WRT posttest scores as the

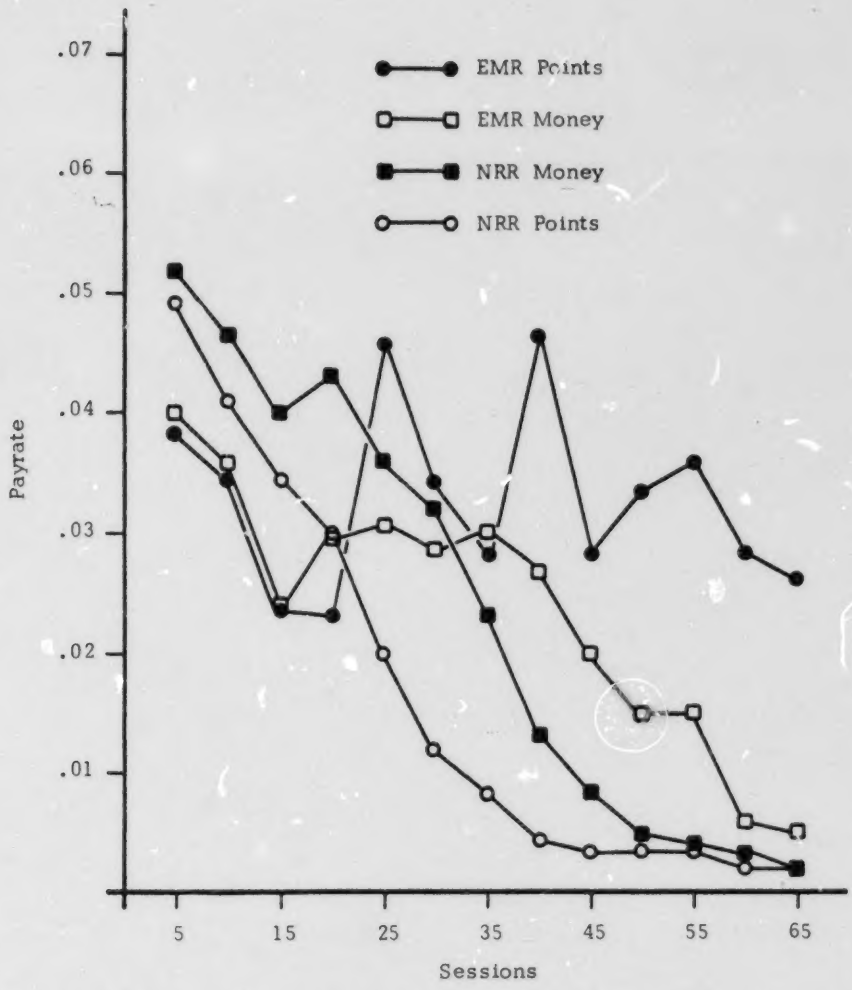

Figure 2. The Payrate (amount of reinforcement including bonuses per reading response) as a Function of Sessions Completed for the Four Experimental Groups. The number .05 on the ordinate equals five one-hundredths of a cent or five-tenths of a point.

criterion and the 100-WRT pretest scores as the covariate are shown. The corresponding group means are presented in Tables 2 and 4. The effect of Treatment was statistically significant, F (1,38) = 4.86, $p < .05$. An inspection of the group mean-gain scores for this variable shows that the difference is in favor of the experimental groups.

The three-way interaction of Type of Subject, Type of Reinforcement, and Treatment on this variable was also statistically significant, F (1,38) = 5.05, $p < .05$. This reflects the fact that the gains evidenced by the various experimental groups and their corresponding control groups were not uniform in magnitude. The mean gains varied from a high of 23.5 for the EMR-Experimental-Money group to a low of 11.25 for the NRR-Control-Points group. Of the four comparisons involving an experimental group and its corresponding control group, three are consistent with the overall treatment effect, whereas one (involving the two EMR-Points groups) is not. The mean difference in gain scores between the experimental groups and their corresponding control groups is 4.85 words.

It should be noted that the effect of Type of Reinforcement was not significant.

The 5-Minute Story Test

All the stories used for the 5-Minute Story Test were from the SRA Reading Laboratory, Ic,

GPO 810-219-4

Table 4

Means and Standard Deviations of Posttest Scores and Gain Scores (GS)[a] by Variable and Groups

	SRA 100-Words		Lorge-Thorndike		California Reading Test		5-Minute Story	
	100-WRT	GS	DIQ	GS	Total Reading	GS	W/Min	% Errors
EMR-$-Exp								
M	44.750	23.500	69.000	4.000	2.850	0.325	48.500	13.850
SD	17.366		9.557		0.387		23.770	10.290
EMR-$-Con								
M	47.500	12.250	69.250	0.750	3.100	0.150	60.600	18.300
SD	29.961		5.439		0.535		37.249	17.847
EMR-Pts-Exp								
M	29.500	14.500	64.250	0.750	2.675	0.250	31.350	17.075
SD	16.299		10.210		0.466		3.873	8.526
EMR-Pts-Con								
M	49.250	16.500	74.750	3.250	3.200	0.350	42.400	12.400
SD	18.118		2.062		0.527		9.812	10.135
NRR-$-Exp								
M	82.125	19.875	80.750	2.625	3.850	0.313	72.375	3.788
SD	9.311		10.236		0.338		18.153	1.985
NRR-$-Con								
M	76.571	18.285	88.000	1.571	4.129	0.415	74.086	5.466
SD	12.594		7.211		0.390		22.563	2.938
NRR-Pts-Exp								
M	75.875	19.625	79.625	-1.500	3.850	0.463	66.900	4.387
SD	11.395		11.795		0.478		18.961	2.043
NRR-Pts-Con								
M	67.875	11.250	79.250	-0.250	3.787	0.187		7.262
SD	15.579		11.068		0.594			4.966

[a] "Gain Score" is the group mean for that measure on the posttest minus the group mean on the corresponding pretest

Table 5

Summary Table for Analysis of Covariance with Posttest DIQ Scores on the Lorge-Thorndike Group Intelligence Test as Criterion and Pretest DIQ Scores as Covariate

Source	df	MS	F
Type of Subject (S)	1	34.304	0.907
Type of Reinforcement (Rf)	1	64.452	1.705
Treatment (T)	1	12.211	0.323
S x Rf	1	29.195	0.772
S x T	1	0.097	0.003
Rf x T	1	12.841	0.340
S x Rf x T	1	34.771	0.920
Within Cells	38	37.810	

Table 6

Analysis of Covariance Summary Table with Posttest California Reading Test, Total Reading, as Criterion and California Reading Test, Total Reading, Pretest as Covariate

Source	df	MS	F
Type of Subject (S)	1	0.203	3.010
Type of Reinforcement (Rf)	1	0.004	0.056
Treatment (T)	1	0.014	0.214
S x Rf	1	0.026	0.391
S x T	1	0.020	0.295
Rf x T	1	0.067	0.993
S x Rf x T	1	0.276	4.089*
Within Cells	38	0.067	

* $p < .05$

Table 7

Analysis of Covariance Summary Table with 100-WRT Posttest Scores as the Criterion Variable and 100-WRT Pretest Scores as the Covariate

Source	df	MS	F
Type of Subject (S)	1	25.289	0.485
Type of Reinforcement (Rf)	1	138.609	2.656
Treatment (T)	1	253.552	4.859*
S x Rf	1	3.433	0.066
S x T	1	5.120	0.098
Rf x T	1	0.231	0.004
S x Rf x T	1	263.594	5.052*
Within Cells	38	52.180	

* $p < .05$

Table 8

Analysis of Covariance Summary Table with Words/Minute Scores from the 5-Minute Story Test as Criterion and Pretest Scores on the 100-WRT as the Covariate

Source	df	MS	F
Type of Subject (S)	1	256.575	0.985
Type of Reinforcement (Rf)	1	596.258	2.288
Treatment (T)	1	3.463	0.013
S x Rf	1	239.956	0.921
S x T	1	50.085	0.192
Rf x T	1	159.934	0.614
S x Rf x T	1	11.622	0.045
Within Cells	38	260.581	

Olive, 2. They were at the same general level of reading ability.

The posttest scores for the experimental and control groups on the 5-Minute Story Test were analyzed using the pretest scores on the 100-WRT Test as the covariate to reduce the variance attributable to initial reading ability as measured by this test. The group means on the 5-Minute Story Test are presented in Table 4. The group means for the pretest scores on the 100-WRT are presented in Table 2. The stories involved in the test were exactly the same for the experiment and control groups. The analysis of the W/Min scores (total number of words read in five minutes divided by five) yielded no significant effects (Table 8). The largest F ratio was for the Type of Reinforcement effect, F (1,38) = 2.288, F = 4.10 needed at .05 level. Thus, the group means did not differ significantly with respect to reading rate at the end of the experiment. (It was interesting to note specifically that because of large individual differences within groups the effect of Type of Subject did not approach significance. There was considerable overlap in the number of words read per minute by the EMR Ss and the NRR Ss.)

An analysis of the %-Errors scores from the 5-Minute Story Test (the total number of errors made in reading for five minutes divided by the total number of reading responses) with the 100-WRT Test pretest scores as the covariate showed that the effect of the Treatment was significant with the experimental groups making significantly fewer errors than the control groups, F (1,38) = 4.524, $p < .05$. The summary of this analysis of covariance is presented in Table 9. The three-way interaction involving the effects of Type of Subject, Type of Reinforcement, and Treatment approached significance, F (1,38) = 3.025, $p < .10$. Of the four comparisons between the experimental groups and the corresponding control groups, three were consistent with the overall treatment effect and favored the experimental groups. The comparison inconsistent with the overall Treatment effect involved the two EMR-Points groups. In

Table 9

Summary Table for Analysis of Covariance with %-Errors on 5-Minute Story Test as Criterion and Pretest Scores on the 100-WRT Test as Covariate

Source	df	MS	F
Type of Subject (S)	1	14.539	0.603
Type of Reinforcement (Rf)	1	15.299	0.635
Treatment (T)	1	109.059	4.524*
S x Rf	1	18.280	0.758
S x T	1	40.399	1.676
Rf x T	1	1.888	0.078
S x Rf x T	1	72.911	3.025
Within Cells	38	24.106	

magnitude the differences in %-Errors performed by the experimental groups and the corresponding control groups averaged 16.9% fewer errors by the experimental groups. If the one reversal of the trend is removed from consideration the average difference is 31.3% fewer errors performed by the experimental groups.

IV
DISCUSSION AND CONCLUSIONS

TEST RESULTS

The Standardized Test Results

No significant differences were found between the experimental and control groups on the Lorge-Thorndike Intelligence Test when the pretest scores were used as a covariate. An analysis of covariance with the California Reading Test, Total Reading, posttest scores as criterion and the pretest scores on the same test as covariate yielded one significant effect. The three-way interaction of Type of Reinforcement, Type of Subject, and Treatment was significant. This effect is due to the differences between the gains of the individual groups on this measure. No meaningful pattern was obvious.

The lack of significant Treatment effects as measured by these two standardized tests is important. Staats et al. (1967) also failed to demonstrate Treatment effects on similar measures. In Staats and Butterfield (1965) the single S involved did show large gains on a standardized reading achievement test.

Several reasons for the lack of significant Treatment effects may be suggested. Staats et al. (1967) suggested three possible explanations: (1) that the training produced effects which did not transfer to the test-taking situation, (2) that the tests used were insensitive dependent variables under these conditions, or (3) that the Ss' test-taking behavior was at fault. The authors favored the third explanation in interpreting their data.

In this study the Ss were tested in small groups with two proctors present. However, it was evident that much guessing still took place. Consistent with Staats et al. (1967), it is suggested that tests such as those used do not constitute adequate measures of Ss' abilities in evaluating research of this type with Ss purposely picked to be at the extremes of the norming populations (if included at all). The effects of such variables as the general motivational states associated with school variables, socioeconomic variables, and test-taking behavior appear stronger than the effects of the program on such tests.

The Nonstandardized Test Results

Two different tests which were devised by the experimenters from SRA materials were administered to the Ss. The first measure (100-WRT Test) was previously used by Staats et al. (1967) and McBurney (1967) in selecting Ss and as a dependent variable. In both studies the mean gain scores of the experimental Ss were greater than those for the control Ss.

In the present study an analysis of covariance with the pretest scores on the 100-WRT Test as the covariate and the posttest scores on the 100-WRT Test as the criterion variable yielded a significant Treatment effect with the experimental Ss showing greater gains. Since the 100 words which formed the test were randomly selected from five of the ten levels of difficulty represented by the materials, it is suggested that each word gained on the test indicates a larger gain in overall reading vocabulary. Although 4253 words were presented in the Individual Word Phase in the total set of materials, none of the Ss finished all of the stories. Thus, the average EMR S either knew or was trained on approximately 1170 different words presented in the program and the average NRR S either knew or was trained on approximately 2070 different words. Within this framework average increases in the number of words responded to correctly on the 100-WRT Test of 105% for the EMR experimental Ss versus 44% for the EMR control Ss constitutes an important gain. The percentage gains for the NRR groups were smaller with the NRR experimental Ss increasing their scores by an average of 33% versus a gain of 25% for the NRR control Ss.

On the 100-WRT Test there was a significant three-way interaction of the Type of Subject,

Type of Reinforcement, and Treatment. The differences among the gains of the individual groups did not present a meaningful pattern.

A second nonstandardized test with high face validity was administered. This test (5-Minute Story Test) consisted of having each S read actual prose materials while the experimenter recorded reading and error rates.

All experimental and control groups were tested at the end of the experiment. The 100-WRT Test pretest scores were used as a covariate to partially remove the variance due to initial reading ability. It was felt that the skills tested in the 100-WRT Test most nearly paralleled those involved in the 5-Minute Story Test.

The analysis of the number of words read per minute (W/Min) revealed no significant differences across groups. An analysis of the percent errors made (%-Errors) showed the effect of the Treatment to be significant. The experimental Ss made significantly fewer errors than did the control Ss. From these data it may be concluded that these remedial reading procedures used for a length of time similar to this study do not significantly affect the reading rate of the Ss. However, the response-contingent reinforcement reduces the number of errors made. Since one would not expect that the reduction in errors would occur on those words not known by the Ss, it seems justifiable to conclude that the procedures significantly reduced the number of careless errors.

Summary of the Test Data

The first hypothesis stated in part that there would be no differences between the experimental and control Ss on the standardized test measures. This was borne out by the data.

The first hypothesis also stated that with respect to nonstandardized tests with high face validity—the 100-WRT Test and the Five-Minute Story Test—significant Treatment effects would be found. This was the case. The experimental treatment resulted in larger gains on these nonstandardized measures for the experimental Ss than were shown by the control Ss. Thus, the first hypothesis was supported.

These results show the nature of some of the gains made by Ss in this remedial reading program. It is concluded that this program increases the word recognition skill of its participants in situations where the words are presented individually. Secondly, there is an increase in the accuracy—of reading words in context as a result of the program.

As to why the increases in word recognition skills and accuracy which resulted from this study did not result in significant gains for the experimental Ss over the control Ss on the Lorge-Thorndike Group Intelligence Test of the California Reading Test, it is suggested that school-related variables such as teacher, facilities, and general motivational states in testing situations are stronger than the effect of the program on this type of test.

The test-taking behavior of the Ss in this study involved considerable random marking. The only tests to show significant differences between the experimental and control groups were those administered in a one-to-one situation. Further evaluations of this program or other similar programs should carefully consider the use of group tests in assessing the status of Ss such as those who have participated in the three studies to date.

Since the basic steps of this program are similar to the basic steps outlined by many other reading specialists (although the details may differ greatly), it is really not surprising that relatively gross measures such as group intelligence tests and reading tests would not show much difference between the experimental and control groups.

The Effect of Type of Reinforcement on the Test Data

The factor of Type of Reinforcement failed to reach statistical significance in any of the analyses of the test data. However, the non-significant differences noted did favor the material reinforcement groups.

The second hypothesis stated that there would be no differences in test performances due to the effect or the Type of Reinforcement factor. This hypothesis was not rejected. The addition of monetary reinforcement to the other reinforcers operating in these procedures (such as the one-to-one situation with an adult, release from the regular classroom, praise, the formal response-contingent awarding of tokens, and the graphs of tokens awarded showing the progress made) did not have a significant effect on those behaviors measured by the tests used.

In general, the response data derived from the performance of Ss in the procedures show that their behavior was adequately maintained. The curves for the mean number of responses made by each experimental group across sessions show that the rate of responding either remains constant, as with the EMR-Points-Experimental group, or increases, as with the other three groups. This is so even though the materials become more difficult as S progresses through the lessons.

DISCUSSION OF DATA INTERNAL TO THE PROGRAM

General Findings

Over the course of the experiment, each S (except one S in the EMR-Money-Experimental group) received 32.5 hours of training. During this period the mean number of responses for the EMR-Points-Experimental group was 24,045. The EMR-Money-Experimental group made an average of over 40,429 reading responses during the same period. The figures for the NRR-Money-Experimental group and the NRR-Points-Experimental group were 57,418 and 62,779, respectively. These figures are below those for Ss in the Staats and Butterfield (1965) study and the Staats et al. (1967) study. In those two studies the mean number of responses ranged from about 65,000 responses made during 40 hours of training by the adolescent S in the first study to an average of over 94,000 responses made by the 18 adolescent Ss during an average of 38.2 hours of training in the second study.

One explanation for the lower total number of reading responses made by the Ss in this study compared to the Ss in the earlier study may be that the 18 Ss in the Staats et al. (1967) study had a higher mean score on the 100-WRT Test than the 24 experimental Ss in this study (69 versus 38.7). This variable was established as a good predictor of Ss' behavior in the program.

Even though the response rate was lower for the Ss participating in this study it is an important demonstration that these younger Ss did make a large number of responses over an extended period of time. Procedures such as the reinforcement system described and used in this study, which permit an extensive investigation of complex behaviors, are an important consideration in the study of educational variables.

As mentioned in the Results section there were two retention tests built into the procedures to measure the retention of the words trained in the Individual Word Phase. On the first of these the rank order of the retention scores of the four experimental groups matched their rank order on the 100-WRT Test pretest scores with the exception of a reversal by the two EMR groups. The range of percentages of words retained was from 69% to 83%. The rank order of the groups on the basis of the long-term retention task scores exactly matched their rank order on the 100-WRT Test pretest scores. The range of these percentages was from 40% to 67%.

From these scores it is clear that the 100-WRT Test predicts Ss performance in a word learning task not only initially (the two alignments were exactly the same on the 100-WRT Test pretest scores and the percentage of correct responses first time through in the Individual Word Phase), but also with respect to relatively long-term retention measures. The percentages retained by all groups showed that considerable learning had taken place during the program. Even though a criterion of one correct, unprompted response was used in training, the EMR Ss retained over a short time span an average of 70.5% of the words missed initially and trained on and 43% of these same words over a longer time (approximately one week). These figures are lower than those reported by Staats et al. (1967) for their lowest group. However, since the average 100-WRT Test score for that group was higher (43.2 as compared to 18.1 for the two EMR groups in this study), one would expect such differences.

The corresponding percentages for the NRR groups are higher. These Ss retained an average of 78.5% of the words trained on in the short-term retention task and an average of 65% of the words on the long-term task. These figures compare favorably with the low group and are quite similar to the middle group in the Staats et al. (1967) study. This is true even though the Ss in this study are younger. The relatively high percentages of words retained after training demonstrate the usefulness of these training procedures in increasing in a permanent fashion the reading vocabulary of low ability Ss.

The third hypothesis stated that there would be no differences with respect to the measures internal to the procedures (such as error rates, number of reading responses made, etc.) between the materially reinforced and nonmaterially reinforced experimental groups. The data do not fully support this hypothesis. An analysis of the total reading responses made using the 100-WRT Test pretest scores as a covariate yielded a significant two-way interaction involving Type of Subject and Type of Reinforcement. The EMR-Money-Experimental group made many more responses than the EMR-Points-Experimental group. The suggestion is strong that the addition of monetary reinforcement to the procedures significantly increased the performance of EMR Ss as measured on this variable.

The NRR groups did not differ to as great an extent. The NRR-Money-Experimental group made fewer responses than the NRR-Points-Experimental group. These data indicate that the addition of monetary reinforcement did not increase the speed of reading with NRR Ss. However, as mentioned earlier, the response-

contingent monetary reinforcement affected the accuracy of Ss. In this program an error resulted not only in having Ss repeat materials, but also in a reduction in the amount earned. Both NRR experimental groups improved in reading accuracy more than their corresponding control groups. However, the amount of improvement was greater for the NRR-Money-Experimental group.

The Amount of Reinforcement Per Reading Response

The data in Figure 2 support in general the earlier finding of Staats and Butterfield (1965) and Staats et al. (1967) that the amount of reinforcement per response decreased as Ss progressed through the program. This result is in part an answer to the frequent question concerning the use of extrinsic reinforcement in training, that is that the S may become dependent upon the reinforcers. It may be suggested that sophisticated methods may be designed to produce the desirable behavior with no lasting dependence upon an artificial system of reinforcement.

The EMR-Points-Experimental group, however, did show some dependence. This exception may be explainable on the basis of the larger percentage of errors made by this group, and/or the more liberal use of bonuses by the instructional technicians administering the program to this group.

The actual amount of total money earned on the average by the EMR-Money-Experimental group Ss was $12.34. This was considerably below the average amount earned by the total group of Ss in the Staats et al. (1967) study.

The totals were not presented for groups in that report so it is not possible to determine whether the low group earned less than the other groups. It is reasonable to assume that this would have been the case if the bonus system had not been used. Because the EMR-Money-Experimental group made fewer responses than the low group in the Staats et al. study it is consistent that they earned a lower amount especially since the bonus system used in this study did not ensure a certain minimal amount earned per day as was the case in the Staats et al. study.

The NRR-Money-Experimental group earned an average of $15.30 over the course of the experiment. This is also lower than the overall average of the Ss in the Staats et al. study. However, as with the EMR Ss, the NRR Ss in this study also made fewer responses on the average than did the Ss of the Staats et al. study. The increases in the total number of reading responses made in each session as the program progressed and relatively low error rates suggest that the differences in the amount earned by the experimental Ss in the two studies is not a critical difference.

BIBLIOGRAPHY

Bijou, S. W. Application of operant principles to the teaching of reading, writing, and arithmetic to retarded children. Address given at the Forty-third Annual Council for Exceptional Children Convention, April, 1965, Portland, Oregon.

Bijou, S. W. Patterns of reinforcement and resistance to extinction in young children. Child Development, 1957, 28, 47-54.

Bijou, S. W., & Sturges, P. T. Positive reinforcers for experimental studies with children— consumables and manipulatables. Child Development, 1959, 30, 151-170.

Birnbrauer, J. S., Bijou, S. W., & Wolf, M. W. A programmed instruction classroom for educable retardates. Paper read at the American Association on Mental Deficiency Convention, May, 1963, Portland, Oregon.

Cole, L. Improvement of reading: With special reference to remedial instruction. New York: Farrar & Rinehart, 1938.

Durrell, D. D. Improvement of basic reading abilities. New York: World Book, 1940.

Fernald, G. M. Remedial techniques in basic school subjects. New York: McGraw-Hill, 1943.

Fernald, G. M., & Keller, H. The effect of kinesthetic factors in the development of word recognition in the case of nonreaders. Journal of Educational Research, 1921, 4, 355-377.

Gates, A. I. The improvement of reading. (Rev. ed.) New York: Macmillan, 1949.

Harris, A. J. How to increase reading ability. New York: Longmans, Green, 1956.

Harrison, M. The story of the initial teaching alphabet. New York: Pitman, 1964.

Hegge, T. G., Kirk, S. A., & Kirk, W. D. Remedial reading drills. Ann Arbor, Michigan: George Wahr, 1945.

Hewett, F. M., Mayhew, D., & Rabb, E. An experimental reading program for neurological impaired, mentally retarded and severely emotionally disturbed children. American Journal of Orthopsychiatry, 1967, 37, 35-48.

Kirk, S. A. Teaching reading to slow-learning children. Boston: Houghton Mifflin, 1940.

McBurney, J. R. The application and extension of learning principles and procedures to the reading remediation of three emotionally disturbed pre-adolescents. Unpublished master's thesis, University of Wisconsin, 1967.

Monroe, M. Children who cannot read. Chicago: University of Chicago Press, 1932.

Otto, W., & McMenemy, R. A. Corrective and remedial teaching. Boston: Houghton Mifflin, 1966.

Schonell, F. J. Backwardness in the basic subjects. London: Oliver & Boyd, 1948.

Staats, A. W. A behavioristic study of verbal and instrumental response hierarchies and their relationship to human problem solving. Unpublished doctoral dissertation, University of California, Los Angeles, 1955.

Staats, A. W. Learning theory and "opposite speech." Journal of Abnormal and Social Psychology, 1957, 55, 268-269. (a)

Staats, A. W. Verbal and instrumental response-hierarchies and their relationship to problem-solving. American Journal of Psychology, 1957, 70, 442-446. (b)

Staats, A. W. Verbal habit-families, concepts, and the operant conditioning of word classes. Psychological Review, 1961, 68, 190-204.

Staats, A. W. A case in and a strategy for the extension of learning principles to the problems of human behavior. In L. Krasner & L. P. Ullman (Eds.), Research in behavior modification. New York: Holt, Rinehart, & Winston, 1965. (a)

Staats, A. W. An integrated-functional learning approach to complex human behavior. Technical Report No. 28 between the Office of Naval Research and Arizona State University 1965. (b) (Reprinted in large part in B. Kleinmuntz (Ed.), Problem

solving: research, method and theory. New York: Wiley, 1966.)

Staats, A. W. Learning, language, and cognition. New York: Holt, Rinehart & Winston, 1968.

Staats, A. W., & Butterfield, W. H. Treatment of nonreading in a culturally-deprived juvenile delinquent: An application of reinforcement principles. Child Development, 1965, 36, 925-942.

Staats, A. W., Finley, J. R., Minke, K. A., & Wolf, M. Reinforcement variables in the control of unit reading responses. Journal of the Experimental Analysis of Behavior, 1964, 7, 139-149.

Staats, A. W., Minke, K. A., Finley, J. R., Wolf, M. M., & Brooks, L. O. A reinforcer system and experimental procedure for the laboratory study of reading acquisition. Child Development, 1964, 35, 209-231.

Staats, A. W., Minke, K. A., Goodwin, W. L., & Landeen, J. "Motivated learning" reading treatment with additional subjects and instructional technicians. Technical Report from the Wisconsin Research and Development Center for Cognitive Learning, University of Wisconsin, 1967, No. 22.

Staats, A. W., & Staats, C. K. Complex human behavior. New York: Holt, Rinehart & Winston, 1963.

Staats, A. W., Staats, C. K., Schutz, R. E., & Wolf, M. The conditioning of reading responses utilizing "extrinsic" reinforcers. Journal of the Experimental Analysis of Behavior, 1962, 5, 33-40.

Staats, A. W., Van Mondfrans, A. P., & Minke, K. A. Manual of administration and recording methods for the Staats "motivated learning" reading procedure. Working Paper from the Wisconsin Research and Development Center for Cognitive Learning, University of Wisconsin, 1967, No. 6.

Witty, P. A., Freeland, A. M., & Grotberg, E. H. The teaching of reading. Boston: D. C. Heath, 1966.

GPO 810-219-2